

## Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720001058
Project Title	Panic Disorder Detection
Maximum Marks	6 Marks

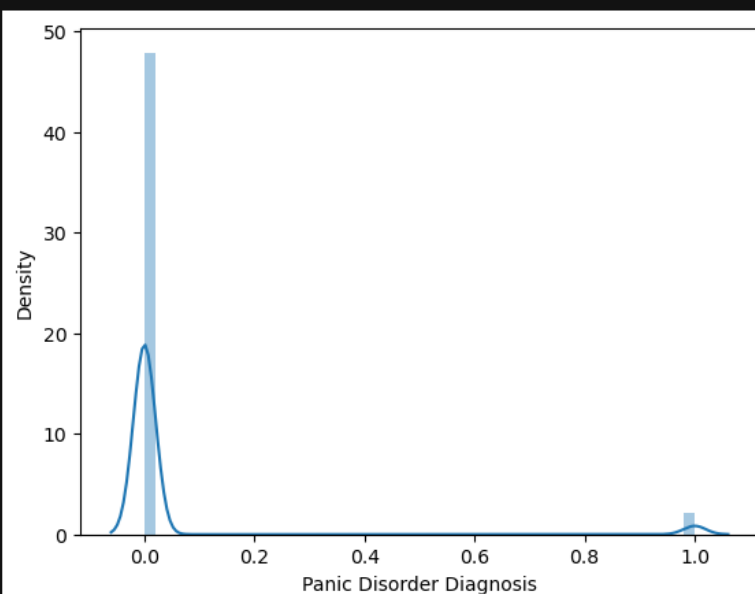
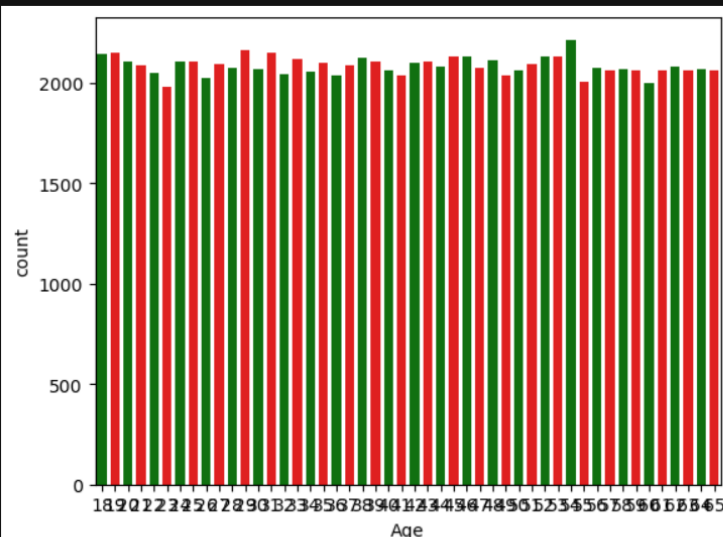
## Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																																																																																																																															
Data Overview	Basic statistics, dimensions, and structure of the data.																																																																																																																																															
	Statistics:																																																																																																																																															
	1.correlation																																																																																																																																															
	<pre>hn = data.corr()</pre>																																																																																																																																															
	<table><thead><tr><th></th><th>Participant ID</th><th>Age</th><th>Gender</th><th>Family History</th><th>Personal History</th><th>Current Stressors</th><th>Symptoms</th><th>Severity</th><th>Social Support</th><th>Lifestyle Factors</th><th>Panic Disorder Diagnosis</th></tr></thead><tbody><tr><th>Participant ID</th><td>1.000000</td><td>-0.001071</td><td>0.001674</td><td>0.003957</td><td>-0.003890</td><td>0.001858</td><td>-0.007720</td><td>-0.001184</td><td>-0.000292</td><td>-0.003263</td><td>0.000182</td></tr><tr><th>Age</th><td>-0.001071</td><td>1.000000</td><td>0.002252</td><td>0.001386</td><td>0.002692</td><td>0.003633</td><td>0.004284</td><td>-0.000413</td><td>0.004726</td><td>0.000854</td><td>-0.000502</td></tr><tr><th>Gender</th><td>0.001674</td><td>0.002252</td><td>1.000000</td><td>0.006519</td><td>0.000724</td><td>0.000553</td><td>-0.000154</td><td>0.001331</td><td>-0.001858</td><td>0.001685</td><td>0.002199</td></tr><tr><th>Family History</th><td>0.003957</td><td>0.001386</td><td>0.006519</td><td>1.000000</td><td>0.002604</td><td>0.000896</td><td>0.002084</td><td>-0.003731</td><td>0.003598</td><td>-0.006489</td><td>0.066927</td></tr><tr><th>Personal History</th><td>-0.003890</td><td>0.002692</td><td>0.000724</td><td>0.002604</td><td>1.000000</td><td>0.002946</td><td>0.003124</td><td>-0.003111</td><td>0.003788</td><td>0.002551</td><td>0.075500</td></tr><tr><th>Current Stressors</th><td>0.001858</td><td>0.003633</td><td>0.000553</td><td>0.000896</td><td>0.002946</td><td>1.000000</td><td>-0.004031</td><td>0.002940</td><td>-0.003106</td><td>-0.003705</td><td>-0.150833</td></tr><tr><th>Symptoms</th><td>-0.007720</td><td>0.004284</td><td>-0.000154</td><td>0.002084</td><td>0.003124</td><td>-0.004031</td><td>1.000000</td><td>-0.004345</td><td>0.003213</td><td>0.003098</td><td>0.043315</td></tr><tr><th>Severity</th><td>-0.001184</td><td>-0.000413</td><td>0.001331</td><td>-0.003731</td><td>-0.003111</td><td>0.002940</td><td>-0.004345</td><td>1.000000</td><td>0.002381</td><td>-0.008878</td><td>0.097926</td></tr><tr><th>Social Support</th><td>-0.000292</td><td>0.004726</td><td>-0.001858</td><td>0.003598</td><td>0.003788</td><td>-0.003106</td><td>0.003213</td><td>0.002381</td><td>1.000000</td><td>-0.004328</td><td>0.003461</td></tr><tr><th>Lifestyle Factors</th><td>-0.003263</td><td>0.000854</td><td>0.001685</td><td>-0.006489</td><td>0.002551</td><td>-0.003705</td><td>0.003098</td><td>-0.008878</td><td>-0.004328</td><td>1.000000</td><td>0.260326</td></tr><tr><th>Panic Disorder Diagnosis</th><td>0.000182</td><td>-0.000502</td><td>0.002199</td><td>0.066927</td><td>0.075500</td><td>-0.150833</td><td>0.043315</td><td>0.097926</td><td>0.003461</td><td>0.260326</td><td>1.000000</td></tr></tbody></table>		Participant ID	Age	Gender	Family History	Personal History	Current Stressors	Symptoms	Severity	Social Support	Lifestyle Factors	Panic Disorder Diagnosis	Participant ID	1.000000	-0.001071	0.001674	0.003957	-0.003890	0.001858	-0.007720	-0.001184	-0.000292	-0.003263	0.000182	Age	-0.001071	1.000000	0.002252	0.001386	0.002692	0.003633	0.004284	-0.000413	0.004726	0.000854	-0.000502	Gender	0.001674	0.002252	1.000000	0.006519	0.000724	0.000553	-0.000154	0.001331	-0.001858	0.001685	0.002199	Family History	0.003957	0.001386	0.006519	1.000000	0.002604	0.000896	0.002084	-0.003731	0.003598	-0.006489	0.066927	Personal History	-0.003890	0.002692	0.000724	0.002604	1.000000	0.002946	0.003124	-0.003111	0.003788	0.002551	0.075500	Current Stressors	0.001858	0.003633	0.000553	0.000896	0.002946	1.000000	-0.004031	0.002940	-0.003106	-0.003705	-0.150833	Symptoms	-0.007720	0.004284	-0.000154	0.002084	0.003124	-0.004031	1.000000	-0.004345	0.003213	0.003098	0.043315	Severity	-0.001184	-0.000413	0.001331	-0.003731	-0.003111	0.002940	-0.004345	1.000000	0.002381	-0.008878	0.097926	Social Support	-0.000292	0.004726	-0.001858	0.003598	0.003788	-0.003106	0.003213	0.002381	1.000000	-0.004328	0.003461	Lifestyle Factors	-0.003263	0.000854	0.001685	-0.006489	0.002551	-0.003705	0.003098	-0.008878	-0.004328	1.000000	0.260326	Panic Disorder Diagnosis	0.000182	-0.000502	0.002199	0.066927	0.075500	-0.150833	0.043315	0.097926	0.003461	0.260326
	Participant ID	Age	Gender	Family History	Personal History	Current Stressors	Symptoms	Severity	Social Support	Lifestyle Factors	Panic Disorder Diagnosis																																																																																																																																					
Participant ID	1.000000	-0.001071	0.001674	0.003957	-0.003890	0.001858	-0.007720	-0.001184	-0.000292	-0.003263	0.000182																																																																																																																																					
Age	-0.001071	1.000000	0.002252	0.001386	0.002692	0.003633	0.004284	-0.000413	0.004726	0.000854	-0.000502																																																																																																																																					
Gender	0.001674	0.002252	1.000000	0.006519	0.000724	0.000553	-0.000154	0.001331	-0.001858	0.001685	0.002199																																																																																																																																					
Family History	0.003957	0.001386	0.006519	1.000000	0.002604	0.000896	0.002084	-0.003731	0.003598	-0.006489	0.066927																																																																																																																																					
Personal History	-0.003890	0.002692	0.000724	0.002604	1.000000	0.002946	0.003124	-0.003111	0.003788	0.002551	0.075500																																																																																																																																					
Current Stressors	0.001858	0.003633	0.000553	0.000896	0.002946	1.000000	-0.004031	0.002940	-0.003106	-0.003705	-0.150833																																																																																																																																					
Symptoms	-0.007720	0.004284	-0.000154	0.002084	0.003124	-0.004031	1.000000	-0.004345	0.003213	0.003098	0.043315																																																																																																																																					
Severity	-0.001184	-0.000413	0.001331	-0.003731	-0.003111	0.002940	-0.004345	1.000000	0.002381	-0.008878	0.097926																																																																																																																																					
Social Support	-0.000292	0.004726	-0.001858	0.003598	0.003788	-0.003106	0.003213	0.002381	1.000000	-0.004328	0.003461																																																																																																																																					
Lifestyle Factors	-0.003263	0.000854	0.001685	-0.006489	0.002551	-0.003705	0.003098	-0.008878	-0.004328	1.000000	0.260326																																																																																																																																					
Panic Disorder Diagnosis	0.000182	-0.000502	0.002199	0.066927	0.075500	-0.150833	0.043315	0.097926	0.003461	0.260326	1.000000																																																																																																																																					
	2. Stats																																																																																																																																															

	<div><pre>data.describe()</pre></div> <table><thead><tr><th></th><th>Participant ID</th><th>Age</th><th>Panic Disorder Diagnosis</th></tr></thead><tbody><tr><td>count</td><td>100000.000000</td><td>100000.000000</td><td>100000.000000</td></tr><tr><td>mean</td><td>50000.500000</td><td>41.454300</td><td>0.04285</td></tr><tr><td>std</td><td>28867.657797</td><td>13.839204</td><td>0.20252</td></tr><tr><td>min</td><td>1.000000</td><td>18.000000</td><td>0.00000</td></tr><tr><td>25%</td><td>25000.750000</td><td>29.000000</td><td>0.00000</td></tr><tr><td>50%</td><td>50000.500000</td><td>41.000000</td><td>0.00000</td></tr><tr><td>75%</td><td>75000.250000</td><td>53.000000</td><td>0.00000</td></tr><tr><td>max</td><td>100000.000000</td><td>65.000000</td><td>1.00000</td></tr></tbody></table> <div>3. Dimensions</div> <div><pre>#showing the dimensions of the data(shape) data.shape #rows = 100000 #columns = 17  (100000, 17)</pre></div>		Participant ID	Age	Panic Disorder Diagnosis	count	100000.000000	100000.000000	100000.000000	mean	50000.500000	41.454300	0.04285	std	28867.657797	13.839204	0.20252	min	1.000000	18.000000	0.00000	25%	25000.750000	29.000000	0.00000	50%	50000.500000	41.000000	0.00000	75%	75000.250000	53.000000	0.00000	max	100000.000000	65.000000	1.00000
	Participant ID	Age	Panic Disorder Diagnosis																																		
count	100000.000000	100000.000000	100000.000000																																		
mean	50000.500000	41.454300	0.04285																																		
std	28867.657797	13.839204	0.20252																																		
min	1.000000	18.000000	0.00000																																		
25%	25000.750000	29.000000	0.00000																																		
50%	50000.500000	41.000000	0.00000																																		
75%	75000.250000	53.000000	0.00000																																		
max	100000.000000	65.000000	1.00000																																		
Univariate Analysis	<div>Exploration of individual variables (mean, median, mode, etc.).</div> <div><table><thead><tr><th></th><th>Participant ID</th><th>Age</th><th>Panic Disorder Diagnosis</th></tr></thead><tbody><tr><td>count</td><td>100000.000000</td><td>100000.000000</td><td>100000.000000</td></tr><tr><td>mean</td><td>50000.500000</td><td>41.454300</td><td>0.04285</td></tr><tr><td>std</td><td>28867.657797</td><td>13.839204</td><td>0.20252</td></tr><tr><td>min</td><td>1.000000</td><td>18.000000</td><td>0.00000</td></tr><tr><td>25%</td><td>25000.750000</td><td>29.000000</td><td>0.00000</td></tr><tr><td>50%</td><td>50000.500000</td><td>41.000000</td><td>0.00000</td></tr><tr><td>75%</td><td>75000.250000</td><td>53.000000</td><td>0.00000</td></tr><tr><td>max</td><td>100000.000000</td><td>65.000000</td><td>1.00000</td></tr></tbody></table></div>		Participant ID	Age	Panic Disorder Diagnosis	count	100000.000000	100000.000000	100000.000000	mean	50000.500000	41.454300	0.04285	std	28867.657797	13.839204	0.20252	min	1.000000	18.000000	0.00000	25%	25000.750000	29.000000	0.00000	50%	50000.500000	41.000000	0.00000	75%	75000.250000	53.000000	0.00000	max	100000.000000	65.000000	1.00000
	Participant ID	Age	Panic Disorder Diagnosis																																		
count	100000.000000	100000.000000	100000.000000																																		
mean	50000.500000	41.454300	0.04285																																		
std	28867.657797	13.839204	0.20252																																		
min	1.000000	18.000000	0.00000																																		
25%	25000.750000	29.000000	0.00000																																		
50%	50000.500000	41.000000	0.00000																																		
75%	75000.250000	53.000000	0.00000																																		
max	100000.000000	65.000000	1.00000																																		
Bivariate Analysis	<div>Relationships between two variables (correlation, scatter plots).</div>																																				

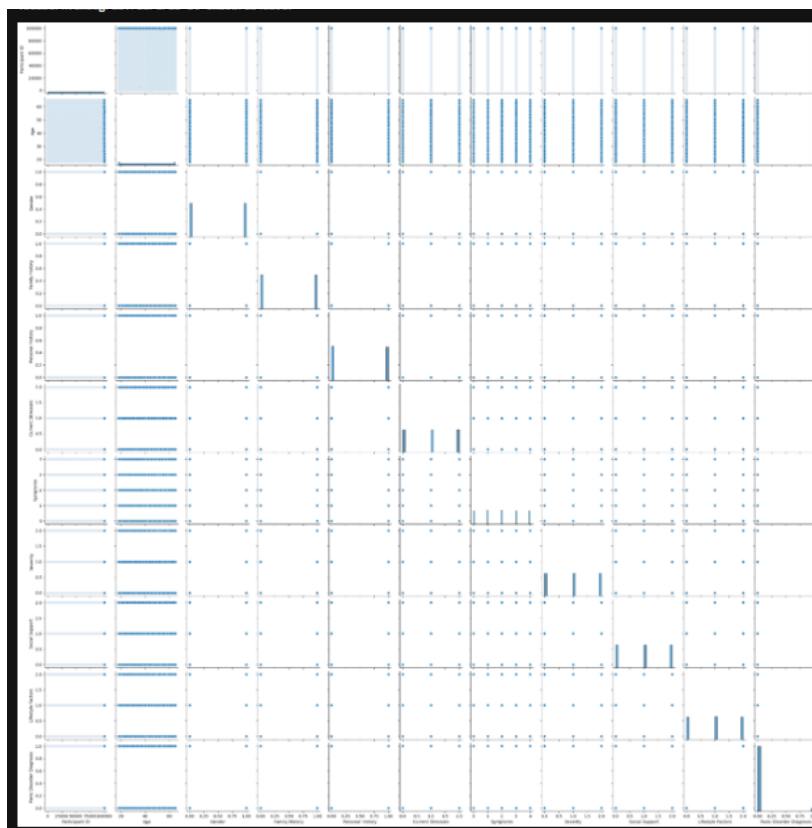
```
sns.countplot(data=data,x='Age',palette = ['green','red'],width = 0.7)
<Axes: xlabel='Age', ylabel='count'>
```



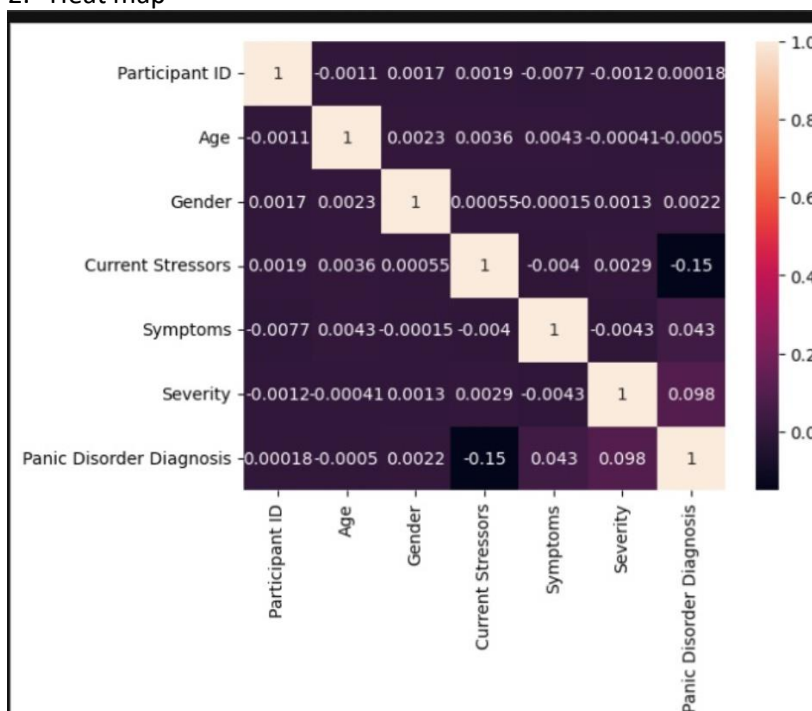
Multivariate Analysis

Patterns and relationships involving multiple variables.  
1.Pair plot

```
sns.pairplot(data)
#multivariate analysis
```



## 2. Heat map



Outliers and Anomalies	Identification and treatment of outliers. No outliers found in box plot																																																																																																																																																																																
Data Preprocessing Code Screenshots																																																																																																																																																																																	
Loading Data	Code to load the dataset into the preferred environment (e.g., Python, R). <div># importing the dataset file data = pd.read_csv('panic_disorder_dataset_training.csv')</div>																																																																																																																																																																																
Handling Missing Data	Code for identifying and handling missing values. 1.identifying missing values <div>#Checking for missing values data.isnull().sum()  Participant ID 0 Age 0 Gender 0 Family History 0 Personal History 0 Current Stressors 0 Symptoms 0 Severity 0 Impact on Life 0 Demographics 0 Medical History 25173 Psychiatric History 24921 Substance Use 33374 Coping Mechanisms 0 Social Support 0 Lifestyle Factors 0 Panic Disorder Diagnosis 0 dtype: int64</div> 3. code to handle missing the value <div>data.drop(columns=['Medical History','Psychiatric History'],inplace = True)  data <table><thead><tr><th></th><th>Participant ID</th><th>Age</th><th>Gender</th><th>Family History</th><th>Personal History</th><th>Current Stressors</th><th>Symptoms</th><th>Severity</th><th>Impact on Life</th><th>Demographics</th><th>Substance Use</th><th>Coping Mechanisms</th><th>Social Support</th><th>Lifestyle Factors</th><th>I Di</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>38</td><td>Male</td><td>No</td><td>Yes</td><td>Moderate</td><td>Shortness of breath</td><td>Mild</td><td>Mild</td><td>Rural</td><td>NaN</td><td>Socializing</td><td>High</td><td>Sleep quality</td><td></td></tr><tr><td>1</td><td>2</td><td>51</td><td>Male</td><td>No</td><td>No</td><td>High</td><td>Panic attacks</td><td>Mild</td><td>Mild</td><td>Urban</td><td>Drugs</td><td>Exercise</td><td>High</td><td>Sleep quality</td><td></td></tr><tr><td>2</td><td>3</td><td>32</td><td>Female</td><td>Yes</td><td>No</td><td>High</td><td>Panic attacks</td><td>Mild</td><td>Significant</td><td>Urban</td><td>NaN</td><td>Seeking therapy</td><td>Moderate</td><td>Exercise</td><td></td></tr><tr><td>3</td><td>4</td><td>64</td><td>Female</td><td>No</td><td>No</td><td>Moderate</td><td>Chest pain</td><td>Moderate</td><td>Moderate</td><td>Rural</td><td>NaN</td><td>Meditation</td><td>High</td><td>Exercise</td><td></td></tr><tr><td>4</td><td>5</td><td>31</td><td>Male</td><td>Yes</td><td>No</td><td>Moderate</td><td>Panic attacks</td><td>Mild</td><td>Moderate</td><td>Rural</td><td>Drugs</td><td>Seeking therapy</td><td>Low</td><td>Sleep quality</td><td></td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>99995</td><td>99996</td><td>22</td><td>Male</td><td>Yes</td><td>No</td><td>High</td><td>Chest pain</td><td>Mild</td><td>Mild</td><td>Rural</td><td>NaN</td><td>Socializing</td><td>Low</td><td>Diet</td><td></td></tr><tr><td>99996</td><td>99997</td><td>57</td><td>Female</td><td>No</td><td>Yes</td><td>Low</td><td>Panic attacks</td><td>Severe</td><td>Mild</td><td>Rural</td><td>NaN</td><td>Meditation</td><td>High</td><td>Diet</td><td></td></tr><tr><td>99997</td><td>99998</td><td>20</td><td>Male</td><td>Yes</td><td>No</td><td>Moderate</td><td>Panic attacks</td><td>Severe</td><td>Moderate</td><td>Rural</td><td>NaN</td><td>Seeking therapy</td><td>Low</td><td>Exercise</td><td></td></tr><tr><td>99998</td><td>99999</td><td>56</td><td>Female</td><td>Yes</td><td>Yes</td><td>High</td><td>Chest pain</td><td>Severe</td><td>Mild</td><td>Rural</td><td>Drugs</td><td>Socializing</td><td>High</td><td>Sleep</td><td></td></tr></tbody></table></div>		Participant ID	Age	Gender	Family History	Personal History	Current Stressors	Symptoms	Severity	Impact on Life	Demographics	Substance Use	Coping Mechanisms	Social Support	Lifestyle Factors	I Di	0	1	38	Male	No	Yes	Moderate	Shortness of breath	Mild	Mild	Rural	NaN	Socializing	High	Sleep quality		1	2	51	Male	No	No	High	Panic attacks	Mild	Mild	Urban	Drugs	Exercise	High	Sleep quality		2	3	32	Female	Yes	No	High	Panic attacks	Mild	Significant	Urban	NaN	Seeking therapy	Moderate	Exercise		3	4	64	Female	No	No	Moderate	Chest pain	Moderate	Moderate	Rural	NaN	Meditation	High	Exercise		4	5	31	Male	Yes	No	Moderate	Panic attacks	Mild	Moderate	Rural	Drugs	Seeking therapy	Low	Sleep quality		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	99995	99996	22	Male	Yes	No	High	Chest pain	Mild	Mild	Rural	NaN	Socializing	Low	Diet		99996	99997	57	Female	No	Yes	Low	Panic attacks	Severe	Mild	Rural	NaN	Meditation	High	Diet		99997	99998	20	Male	Yes	No	Moderate	Panic attacks	Severe	Moderate	Rural	NaN	Seeking therapy	Low	Exercise		99998	99999	56	Female	Yes	Yes	High	Chest pain	Severe	Mild	Rural	Drugs	Socializing	High	Sleep	
	Participant ID	Age	Gender	Family History	Personal History	Current Stressors	Symptoms	Severity	Impact on Life	Demographics	Substance Use	Coping Mechanisms	Social Support	Lifestyle Factors	I Di																																																																																																																																																																		
0	1	38	Male	No	Yes	Moderate	Shortness of breath	Mild	Mild	Rural	NaN	Socializing	High	Sleep quality																																																																																																																																																																			
1	2	51	Male	No	No	High	Panic attacks	Mild	Mild	Urban	Drugs	Exercise	High	Sleep quality																																																																																																																																																																			
2	3	32	Female	Yes	No	High	Panic attacks	Mild	Significant	Urban	NaN	Seeking therapy	Moderate	Exercise																																																																																																																																																																			
3	4	64	Female	No	No	Moderate	Chest pain	Moderate	Moderate	Rural	NaN	Meditation	High	Exercise																																																																																																																																																																			
4	5	31	Male	Yes	No	Moderate	Panic attacks	Mild	Moderate	Rural	Drugs	Seeking therapy	Low	Sleep quality																																																																																																																																																																			
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...																																																																																																																																																																		
99995	99996	22	Male	Yes	No	High	Chest pain	Mild	Mild	Rural	NaN	Socializing	Low	Diet																																																																																																																																																																			
99996	99997	57	Female	No	Yes	Low	Panic attacks	Severe	Mild	Rural	NaN	Meditation	High	Diet																																																																																																																																																																			
99997	99998	20	Male	Yes	No	Moderate	Panic attacks	Severe	Moderate	Rural	NaN	Seeking therapy	Low	Exercise																																																																																																																																																																			
99998	99999	56	Female	Yes	Yes	High	Chest pain	Severe	Mild	Rural	Drugs	Socializing	High	Sleep																																																																																																																																																																			
Data Transformation	Code for transforming variables (scaling, normalization). <div>from sklearn.preprocessing import LabelEncoder label_encoder = LabelEncoder()  #finding the categorical data and converting them into numeriv by label encoding technique columns_to_encode = ['Gender', 'Family History', 'Personal History', 'Current Stressors', 'Symptoms', 'Severity', 'Impact on Life', 'Demographics', 'Substance Use', 'Coping Mechanisms', 'Social Support', 'Lifestyle Factors']  # Encode each column for col in columns_to_encode:     if data[col].dtype == 'object':         data[col] = label_encoder.fit_transform(data[col].astype(str))  # Display the transformed DataFrame data.head()</div>																																																																																																																																																																																

Feature Engineering	Code for creating new features or modifying existing ones. No new features were created
Save Processed Data	Code to save the cleaned and processed data for future use. <pre>import pickle pickle.dump(sc,open("panic.pkl",'wb'))</pre>